

Genome Sequence of the Hemolytic-Uremic Syndrome-Causing Strain *Escherichia coli* NCCP15647

Haeyoung Jeong,^{a,c} Fumei Zhao,^{b,c} Davaajargal Igori,^{b,c} Kyung-Hwan Oh,^d Seon-Young Kim,^{e,f} Sung Gyun Kang,^{g,h} Byung Kwon Kim,^a Soon-Kyeong Kwon,^{a,c} Choong Hoon Lee,^a Ju Yeon Song,^a Dong Su Yu,^a Mi-Sun Park,^d Seung-Hak Cho,^d and Jihyun F. Kim^{a,i}

Systems and Synthetic Biology Research Center,^a Green Bio Research Center,^b and BioMedical Genomics Research Center,^c Korea Research Institute of Bioscience and Biotechnology (KRIBB), Yuseong-gu, Daejeon, Republic of Korea; Biosystems and Bioengineering Program,^e Department of Functional Genomics,^f and Department of Marine Biotechnology,^g University of Science and Technology, Yuseong-gu, Daejeon, Republic of Korea; Division of Enteric Bacterial Infections, Center for Infectious Diseases, National Institute of Health, Osong Health Technology Administration Complex, Gangoe-myeon, Cheongwon-gun, Chungcheongbuk-do, Republic of Korea^d; Korea Ocean Research and Development Institute, Sangnok-gu, Ansan, Republic of Korea^h; and Department of Systems Biology, Yonsei University, Seodaemun-gu, Seoul, Republic of Koreaⁱ

Enterohemorrhagic *Escherichia coli* (EHEC) causes a disease involving diarrhea, hemorrhagic colitis, and hemolytic-uremic syndrome (HUS). Here we present the draft genome sequence of NCCP15647, an EHEC isolate from an HUS patient. Its genome exhibits features of EHEC, such as genes for verotoxins, a type III secretion system, and prophages.

Infection by enterohemorrhagic *Escherichia coli* (EHEC) that produces Shiga toxin or verotoxin can cause severe watery or bloody diarrhea and can be fatal for children and the elderly (12, 17). Although *E. coli* O157:H7 is the main EHEC serotype (10, 14), non-O157 EHEC strains such as O26, O111, or O103 have been recently reported (4, 7, 8). In addition, genome sequence analysis revealed that the O104:H4 outbreak in Germany was caused by an EHEC strain that had acquired virulence genes from another diarrheagenic *E. coli* strain (5, 13). EHEC infection in Korea was first documented in 1998 (11), and there were two outbreaks in 2003 and 2004 (6). Since then, several dozens of cases have been reported yearly. To trace back the possible occurrence of EHEC infection in Korea and to gain insights into EHEC pathogenic potential and genome organization, we determined the genome sequence of the verotoxin-producing strain NCCP15647, which had been isolated in 2003 from a Korean male patient who suffered from hemolytic-uremic syndrome.

The draft genome sequence of NCCP15647 was obtained by a whole-genome shotgun strategy. Over 300-fold Illumina GA Iix paired-end reads produced from a 550-bp library were *de novo* assembled using the CLC Genomics Workbench (CLC bio, Inc.). Automatic scaffolding and gap closing using the 231 contigs and the same paired-end data set were performed by SSPACE (3) and IMAGE (16), respectively. The final assembly contains 5,146,299 bp in 126 contigs over 200 bp (133,662-bp N_{50} and 279,619-bp maximum contig size) with 51.61% G+C content. Additional scaffolding using 3-kb mate-pair library reads (40-nucleotide cycle, 26,721,171 read pairs) yielded 73 scaffolds with a maximum size of 1.51 Mb.

Contig sequences obtained from automatic gap closing were used for genome annotation using the RAST server (1). A total of 68% of 5,055 coding sequences were assigned into 603 subsystem categories. A multilocus-based phylogenetic tree obtained using 14 genes of completely sequenced *E. coli* genomes revealed that NCCP15647 belongs to the B1 phylogroup, the enterotoxigenic strain E24377 (O139:H28) being the closest one. When average nucleotide identity was used (15), however, the commensal strain IAI1 (O8) was closer to NCCP15647 (99.22% versus 98.97% by MUMmer). Sequence analysis using the *wzx*, *wzy*, and *fliC* genes (2) indicated that *fliC* is the most discriminating gene.

The genome contains two genes for verotoxins I and II. Contrary to the common integration of the locus of enterocyte effacement (LEE) pathogenicity island harboring a type III secretion system (TTSS) and *eae* genes at the *selC* site in other typical EHEC strains, such as O157:H7, the TTSS from NCCP15647 was found at another site without *eae* and was not associated with prophages. We could also identify several prophage-related gene clusters and at least three putative plasmid contigs harboring genes for replication, one of them also having a colicin gene. Genome sequence analysis can rapidly reveal the genetic makeup of the causative agents of a serious infection, which can help source tracking of infection and proper treatment.

Nucleotide sequence accession number. The draft sequence determined in this study has been deposited in GenBank under the accession no. [AJMB000000000](http://www.ncbi.nlm.nih.gov/nuccore/AJMB000000000). The sequence and annotation are also available from the Genome Encyclopedia of Microbes (GEM) site at <http://www.gem.re.kr/> (9).

ACKNOWLEDGMENTS

This project was conducted in part as the Microbial Genome Sequencing and Analysis Course at the University of Science and Technology in the fall of 2011. We are grateful to Seong Keun Kim, Daehee Lee, and students in the class for critical comments on the manuscript and members of GEM and the KRIBB sequencing team for technical help.

This work was supported by the National Research Foundation of Korea (2011-0017670 to J.F.K.) of the Ministry of Education, Science and Technology, the KRIBB Research Initiative Program, and the National Institute of Health (NIH 4800-4845-300 to S.H.C.), Republic of Korea.

REFERENCES

1. Aziz RK, et al. 2008. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9:75. doi:10.1186/1471-2164-9-75.

Received 18 April 2012 Accepted 30 April 2012

Address correspondence to Jihyun F. Kim, jfk1@yonsei.ac.kr, or Seung-Hak Cho, skcho38@korea.kr.

H.J., F.Z., and D.I. contributed equally to this article.

Copyright © 2012, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JB.00651-12

2. Ballmer K, et al. 2007. Fast DNA serotyping of *Escherichia coli* by use of an oligonucleotide microarray. *J. Clin. Microbiol.* **45**:370–379.
3. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**:578–579.
4. Brooks JT, et al. 2005. Non-O157 Shiga toxin-producing *Escherichia coli* infections in the United States, 1983–2002. *J. Infect. Dis.* **192**:1422–1429.
5. Chattaway MA, Dallman T, Okeke IN, Wain J. 2011. Enteroaggregative *E. coli* O104 from an outbreak of HUS in Germany 2011, could it happen again? *J. Infect. Dev. Ctries.* **5**:425–436.
6. Cho H-S, Cho M-C, Noh S, Kim M-N, Kim K-M. 2010. A case of verotoxin-producing *Escherichia coli* O157:H7 with hemorrhagic colitis in an infant, diagnosed by multiplex PCR. *Korean J. Clin. Microbiol.* **13**:85–89.
7. Feng PC, et al. 2010. Prevalence, characterization and clonal analysis of *Escherichia coli* O157: non-H7 serotypes that carry *eae* alleles. *FEMS Microbiol. Lett.* **308**:62–67.
8. Gould LH, et al. 2009. Recommendations for diagnosis of Shiga toxin-producing *Escherichia coli* infections by clinical laboratories. *MMWR Recomm. Rep.* **58**:1–14.
9. Jeong H, Yoon SH, Yu D-S, Oh T-K, Kim JF. 2008. Recent progress of microbial genome projects in Korea. *Biotechnol. J.* **3**:601–611.
10. Kaper JB. 1998. Enterohemorrhagic *Escherichia coli*. *Curr. Opin. Microbiol.* **1**:103–108.
11. Kim YB, et al. 1998. Isolation of an *Escherichia coli* O157:H7 strain producing Shiga toxin 1 but not Shiga toxin 2 from a patient with hemolytic uremic syndrome in Korea. *FEMS Microbiol. Lett.* **166**:43–48.
12. Levine MM. 1987. *Escherichia coli* that cause diarrhea: enterotoxigenic, enteropathogenic, enteroinvasive, enterohemorrhagic, and enteroadherent. *J. Infect. Dis.* **155**:377–389.
13. Mellmann A, et al. 2011. Prospective genomic characterization of the German enterohemorrhagic *Escherichia coli* O104:H4 outbreak by rapid next generation sequencing technology. *PLoS One* **6**:e22751. doi:10.1371/journal.pone.0022751.
14. Pennington H. 2010. *Escherichia coli* O157. *Lancet* **376**:1428–1435.
15. Richter M, Rossello-Mora R. 2009. Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. U. S. A.* **106**:19126–19131.
16. Tsai IJ, Otto TD, Berriman M. 2010. Improving draft assemblies by iterative mapping and assembly of short reads to eliminate gaps. *Genome Biol.* **11**:R41. doi:10.1186/gb-2010-11-4-r41.
17. Welinder-Olsson C, Kaijser B. 2005. Enterohemorrhagic *Escherichia coli* (EHEC). *Scand. J. Infect. Dis.* **37**:405–416.